International Journal of Computer Engineering and Sciences Research

VOL. 06, NO. 01, September-October 2024
Pages 01–17
Available online at: www.ijcesr.com

A Method of Hybridizing and Cascading Several Learning Phases to Detect Brain Stroke in MRI and CT Scan Images

Alhina Afrin Dolon,A. H. M. Kamal

Department of Computer Science & Engineering,*JatiyaKabiKaziNazrul Islam University, Namapara, Mymensingh, 2200, Mymensingh, Bangladesh*

Nafiz Imtiaz Rafin,Md. Golam Rabiul Alam

Department of Computer Science & Engineering,
BRAC University, 66 Mohakhali, Dhaka, 1212, Dhaka, Bangladesh

M. Ali AkberDewan

School of Computing and Information Systems, Athabasca University, Canada

Abstract: Brain stroke is a fatal and significant health issue nowadays. Machine learning (ML) models offer a rapid and precise prediction result and have developed into a potent tool in healthcare settings, providing brain stroke patients with individualized therapeutic treatment. Although there have been some studies on predicting brain stroke, these studies fail to offer flawless predictions and there are always some miss-classifications. Flawless prediction of brain stroke is crucial because the error in prediction could be harmful to the patients and disrupt the patient's proper medical care. Thus, in this study, we propose an ensemble architecture consisting of seven different machinelearning models. The proposed system was evaluated on a bench-marking dataset, where the system showed competitive performance.

Keywords*—Brain Stroke; MRI; CTI Scan; XGBoost; Ensemble;Convolutional Neural Network*

I. INTRODUCTION

The number of different cerebral vascular disorders rises globally[1] and these illnesses hold the top spot. When brain cells do not receive enough oxygen, damage to the brain cells begins, a condition known as cerebral vascular disease occurs. Cardiovascular diseases were among the top ten global fatalities, according to data recently provided by the World Health Organization WHO) [2]. The primary classifications of cerebrovascular illnesses include aneurysms, ischemic strokes, and hemorrhagic strokes. Vascular narrowing, obstruction, and bleeding are the causes of strokes because they obstruct the oxygen supply to the brain's cells.

Stroke ranks as the second leading cause of death and disability globally[3]. A stroke, sometimes called a cerebrovascular accident, is a neurological illness that is characterized by ischemia or bleeding of the brain's arteries. It usually causes a variety of functionally harmful motor and cognitive abnormalities. Stroke is one of the main causes of disability in both developed and developing countries.According to the 2016 Global Burden of Disease, Injuries and Risk Factors Analysis, strokes caused 116 million years of life with a disability and 5.5 million fatalities each year[4]. The American Heart Association considers stroke to be a severe health issue due to its high death rate[5].In addition, as the costs of hospitalizing stroke patient's rise, there has been a greater need for advanced technologies to support clinical diagnosis, treatment, clinical event forecasting, therapeutic approach suggestions, rehabilitation programs, etc.[6].Since a stroke cannot be contracted from another person, it can be prevented by increasing knowledge and managing risk factors such as diabetes, hypertension, dyslipidemia, and atrial fibrillation. Furthermore, patients whose prognosis after a stroke is predicted to be poor may benefit from intensive care. Therefore, the capacity to predict stroke outcome is critical for both promptly initiating therapy and allocating medical resources.

Several research on the enhancement of stroke diagnosis using ML in terms of accuracy and speed have been undertaken over the last several decades. The contributions of this study are summarized as follows: (1) A novel ensemble technique has been presented that suppresses the performance of a series of machine learning model; (2) precise classification of brain strokes and normal cases; and (3) XAI.

A concise synopsis of prior research on several frameworks for anomaly identification is provided in Section 2 of this article. Section 3 provides a detailed breakdown of our framework. This section of consisted of our System Model, Data Description, and Model Specification sections. Section 4.2 presents the performance measures we used and the outcomes we attained after putting our system through its paces. We reach the conclusion of our article in Section 5.

II. RELATED WORKS

Researchers are working on ways to avoid this condition and develop an effective remedy due to the rising number of patients being affected with brain strokes and the significant expenses associated with it. As a result, numerous studies utilizing different approaches have been conducted recently on MRI and CT scan images. Moreover, it has become increasingly common to employ neuroimaging to assess a variety of neurological conditions.

In order to determine an irregularity index and correlate it with individuals with Alzheimer's disease (AD) and Alzheimer's disease (MCI), Oliveira et al.[7]Evaluated an unsupervised v-One-Class Support Vector Machine (- OC-SVM) trained with neuroimaging factors, such as cortical thickness and cerebral volume, from healthy individuals. With an accuracy of 84.3%, the method effectively detected outliers among AD patients.

By using brain magnetic resonance imaging (MRI) assessments of the left temporal lobes, bilateral dorsolateral prefrontal regions, and left medial parietal lobes, Schizophrenia, et al. were able to distinguish between patients with childhood-onset schizophrenia and healthy individuals[8]. The method effectively identified classes with 73.7% accuracy, and a larger brain-based potential of disease was connected with both statistically significant lower functioning and less developmental delays. Machine learning may also be used to distinguish between various disease subtypes. Bleich-Cohen et al. [9] employed Searchlight Based Feature Extraction (SBFE), a data-driven multi-voxel pattern analysis (MVPA) technique, to look for clusters of cognitive stress response in brain functional magnetic resonance imaging (fMRI). This machine learning (ML) method effectively discriminated with 91% accuracy between the two subgroups of schizophrenia patients with and without Obsessive-Compulsive Disorder (OCD) in order to distinguish between symptom intensity and a psychiatric comorbidity.

An et al. [10] evaluated whole white matter abnormalities between patients with mesial temporal epilepsy and matched normal participants using a machine learning technique that assessed tract-based spatial statistics, including fractional anisotropy. This ML-based method correctly distinguished between each group and shown high sensitivity to changes in fractional anisotropy in people with mesial temporal epilepsy, despite the fact that no lesion could be found on neuroimaging.Moghim et al. suggested a prediction model for the incidence of seizures in a single patient[11]. This approach was based on a multi-class support vector machine (SVM) and 14 carefully selected features of an electroencephalogram in epilepsy patients. With a window of 20 to 25 minutes, average sensitivity of 90.15, specificity of 99.44%, and accuracy of 97% were recorded for the expected time of seizure.

Determining the affected areas, the extent of the damage, and subsequently the functional outcome is made easier by estimating the burden of lesions in patients suffering from multiple sclerosis, Alzheimer's disease (AD), traumatic brain injury (TBI), and dementia. Lesion segmentation using a multi-modal brain MRI was proposed by Kaminatas et al.[12], based on an 11-layer deep, multi-scale, 3D Convolutional Neural Networks (CNN) called Deep Medic. Their proposed unique training technique consists of a 3D CNN that creates proper soft segmentation maps and a linked Conditional Random Field that imposes regularization requirements on the CNN output to produce the final hard segmentation labels. This allows for a deeper and considerably more discriminative delimitation of lesion load, with the best reported accuracy observed in a cohort of patients with severe traumatic brain injury.

According to Lin et al.[13], an ML technique incorporating 206 clinical factors could predict the functional prognosis of patients with ischemic and hemorrhagic stroke 90 days later with an area under the receiver operating characteristics (AUROC) of 0.946 to 0.970. They evaluated several supervised machine learning algorithms' ability to forecast 90-day mRS outcomes using data from a national stroke database. The results indicated that a strong tool for predicting stroke outcomes may be created by using machine learning techniques on a large dataset for feature selection and training and exposure. The follow-up information is particularly helpful for predicting outcomes. If two classifiers have similar performance, we may select the prediction model based on real-world needs and how well it performs with patients who have had severe strokes.

In order to anticipate the poor functional prognosis in patients with acute ischemic stroke, Heo et al. observed that DL algorithms outperformed the Acute Stroke Registry and Analysis of Lausanne (ASTRAL) score, a commonly used logistic regression-based method[14]. (AIS). When forecasting the prognosis of stroke patients utilizing statistical data, DL algorithms performed noticeably better than did conventional statistical models. Additionally, the accuracy of DL systems employing brain magnetic resonance imaging (MRI) in forecasting the eventual infarct size and reperfusion status increased.

For the purpose of predicting the aggregate functions in AIS patients who underwent endovascular therapy, Hilbert et al.[15]Claimed that computed tomography angiography DL employing ResNet and an auto encoder may provide effective image biomarkers.

III. Methodology

This section has discussed our extensive technique, which is divided into three parts. Following the discussion of our overall system model in subsection 3.1, we address the data collection and processing in subsection 3.2. Finally, we provide our model specifications in subsection 3.3. In theFigure 1, the proposed model of our research is illustrated.

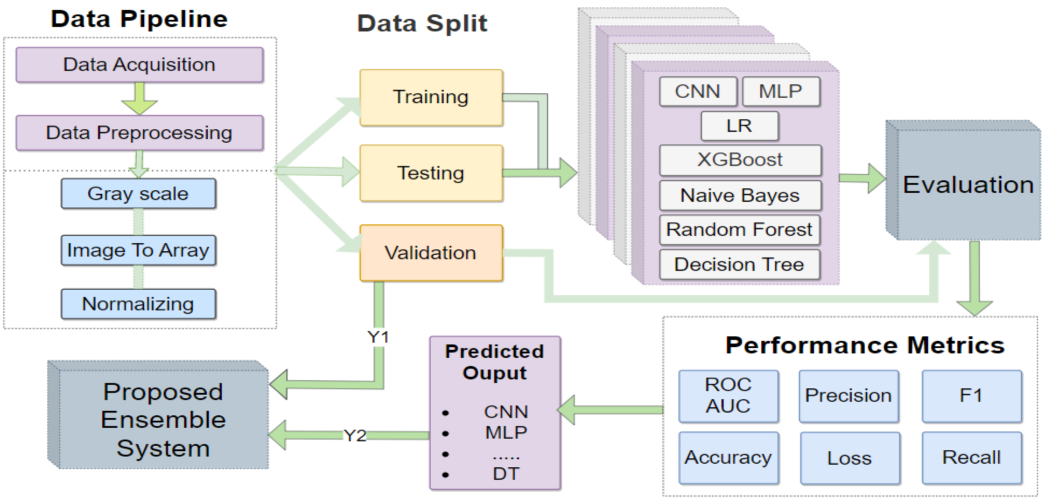

Figure 1: Proposed System Model.

1.1 System Model

Initially, after data acquisition we employed a series of data preprocessing techniques. First, we converted our images to gray scale and then to arrays. Finally, we normalized the data and move on to data splitting section where we split our data into three separate sections. Training and testing is employed while training and validation is utilized while testing the trainedmodel. We have used both brain magnetic resonance imaging (MRI) and Computed Tomography (CT) Scan images separately to run the project. We have employed seven different machine learning model that are Convolution Neural Network,K-Nearest Neighbors Classifier, Logistic Regression, XGBoost, MLP, Naive Bayes and Decision Tree Classifier on MRI images. And We have employed seven different machine learning model that are Convolution Neural Network, MLP, Logistic Regression, XGBoost, Naive Bayes, Random Forest Classifier and Decision Tree Classifier on CT Scan images.A number of performance indicators, including accuracy, recall,loss,f1-score, precision, AUC ROC and confusion matrix, were taken into consideration throughout the evaluation stage. With the validation data we test our model against each of the trained model and store its predictions into a array. This array becomes our X for our system's next stage. And y is the true y values of the validation set. We than merge validation y true values with our newly created array which is X and shuffle it. For the next stage we further divide our newly created dataset into two section. Finally we train a new model and test its score against the test data.

1.2 Data Acquisition and Preparationnits

1.2.1 Data Acquisition and Description

For our research, the first selection of dataset we have used for MRI image is the Acute Ischemic Stroke MRI image dataset[16] which consists of Diffusion MRI image collected and ethically approved by TurgutOzalUni- ¨ varsity medical faculty hospital neurology department. For CT scan image, we have used Brain Stroke Prediction CT Scan Image Dataset[17] which is an open source dataset from kaggle. Figure 2presents the sample of brain CT scan images from our acquired dataset. Here, (a) represents normal or no strokes and (b) represents strokes. Moreover,Figure 3 for CT scan images depicts our data distribution. We have 6650 cases of normal CT scan images and 10857 instances of abnormal or stroke CT images. InFigure 4 for MRI images depicts data distribution, we have 8718 MRI images data where 4520 cases of normal MRI images and 4198 cases are abnormal or stroke MRI images.

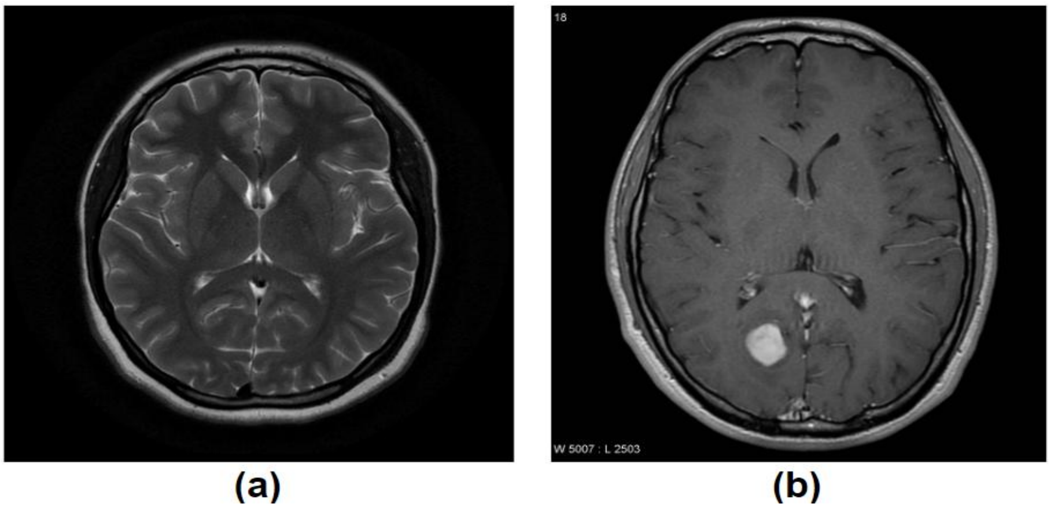

Figure 2: Brain Stroke CT Scans Sample. Here,(a) represents normal or no stroke and (b) depicts strokes.

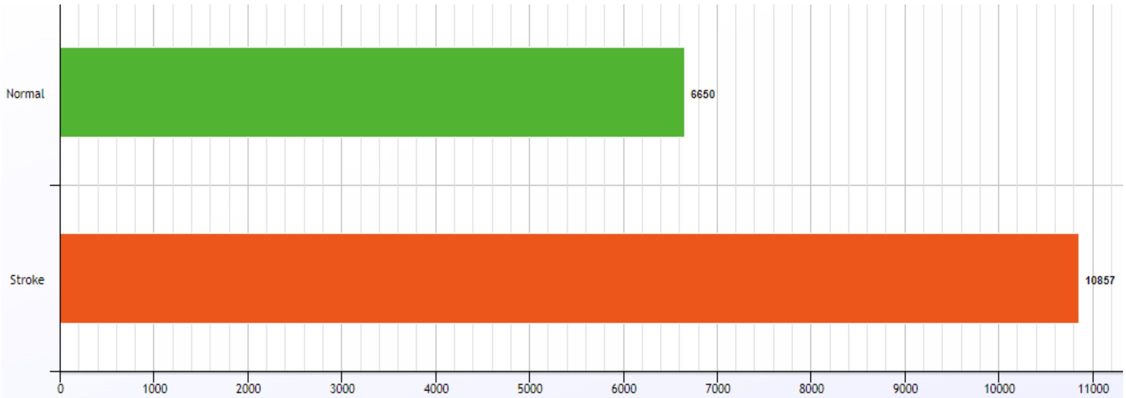

Figure 3: Brain Stroke CT Scans Data Quantity. Sample. Here, green represents normal or no stroke and red depicts strokes.

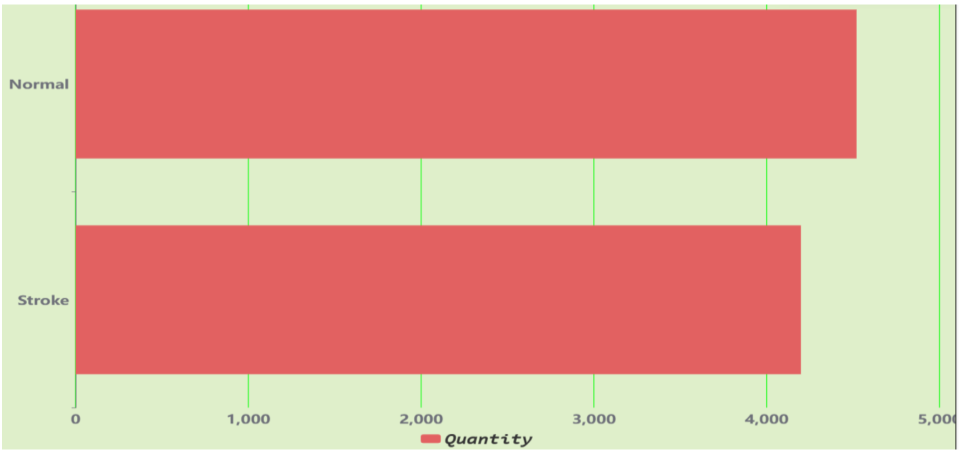

Figure 4: Brain Stroke MRI Images Data Quantity. Sample

1.2.2 Data Preparation

In this article, brain strokes have been classified utilizing augmentation methods outlined in the article. We assessed a wide range of data augmentation methods that have since been used in previous studies. Our major concerns are to understand the consequences of data augmentation and figuring out if one sort augmentation is better than another. We finally settled on our data augmentation strategy, which turned out to be quite effective, after conducting a thorough research on previously used augmentation techniques and conducting a few trial runs.

The techniques we utilized are presented below:

- Rotate
- Horizontal Flipping
- Vertical Flipping
- Left Shifting
- Right Shifting
- Translating
- Resizing
- Noise Injection

After performing data augmentation we move on to our next step that is gray scale conversion. Later we convert the images into arrays and finally this array is normalized before feeding it to our machine learning algorithm.

1.3 Model Specification

1.3.1 Random Forest Classifier

Random forest algorithm is a mixture of tree classifiers in which a random vector is sampled independently for every class and a voting unit is selected to allocate the input vector for the general category for every one of the tree classes [18]. Finally, RF selects the best choice by voting. It is superior to a decision-making tree since it eliminates surplus by integrating the results. To develop the model, arbitrary data from the data set is used, and decision trees are built depending on every sample data. After every decision tree produces a classification results, voting is undertaken, and the prediction by the most votes is allocated to that specific data. Our RF model employs the Gini index, which is among the most widely adopted metrics for determining the attribute's legitimacy in relation to the classes which can be written as Equation 1.

$$G = \sum_{i=1}^{C} p\,(i) * \left(1 - p(i)\right) \quad (1)$$

Where,
pi= Probability of considering data point of class i.

1.3.2 Extreme Gradient Boosting

Decision trees are used in the gradient boosting ensemble machine learning technique known as XGBoost [19]. Because of its great accuracy and effectiveness, it is frequently employed in many different fields. By approximating the second order derivative, XGBoost aims to reduce the inaccuracy of the overall model. It has been demonstrated that adaptability is attributable to several significant frameworks and algorithm advancements, including a novel tree supervised learning, in all XGBoost scenarios.

The use of variables to dynamically alter the classifier and enhance its precision or understandability is one of its most useful features. To begin with, the optimization techniques of XGBoost contribute to the learner's improved comprehensibility. The second most helpful aspect of XGBoost is the incorporation of regularization methods that reduce model bias, such as Ridge and Lasso regression. Regularization is essentially a method of reducing variance and sampling error by adjusting the expected coefficients. It has been suggested that regularization, sometimes referred to as regularized regression, can be used to do generalization, early stopping, sparsity manipulation, etc. The XGBoost formula is represented by equation 2.

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, y_i^{\wedge(t-1)} + f_t(X_i)\right) + \Omega(f_t) \quad (2)$$

1.3.3 Gaussian Naive Bayes

A Naive Bayes variation that fits the Gaussian normal distribution and handles continuous input is called Gaussian Naive Bayes. Based on a Bayesian network, it uses an easily modifiable technique to make predictions quickly and in real time. This method is usually applied to real-world problems since it can be quickly configured to respond to user requests. It is typical to assume that each group's continuous values are distributed when working with continuous data. The likelihood of the features is presumed in Equation 3.

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right) \quad (3)$$

1.3.4 Decision Tree Classifier

The decision tree is one of the most popular and useful techniques for forecasting and classification.It's a kind of tree structure that looks like a roadmap, with a test on an attribute at each node, a test result at each branch, and a class label at each child node. Decision trees classify events by placing them in a tree structure that assigns a category to each instance as it progresses from the root to a leaf node. Before moving down the branch of the tree that corresponds to the attribute's value, one must test the attribute provided by the tree's root node in order to classify an instance. After then, the sub-tree rooted at the new node goes through the same process again. The decision tree in the above diagram assigns a classification to each leaf based on whether a particular morning is good for playing tennis and returns that classification.

Data at node m can be represented by Q_m with instance n_m. For each $\theta = (j, t_m)$ intance split consisting of a feature j and threshold t_m, separate the data into $Q_m^{left}(\theta)$ and $Q_m^{right}(\theta)$ subsets

$$\begin{aligned} Q_m^{left}(\theta) &= \{(x,y) | x_j \leq t_m\} \\ Q_m^{right}(\theta) &= Q_m \backslash Q_m^{left}(\theta) \end{aligned} \tag{4}$$

Then, depending on the job at hand, either a loss function or an impurity function is used to calculate the grade of a potential division of node m.

1.3.5 Multilayer Perceptron (MLP)

The multi-layer perceptron (MLP) is a supplement to feed-forward neural networks [20]. There are three distinct types of layers in it: hidden units, output units, and input nodes. The input layer receives the signal that has to be analyzed. The output layer performs the required functions, including classification and prediction. This model’s actual computing engine is made up of an endless number of hidden layers nested between the input and output levels. This algorithm’s input to output layer transfers data in a way that is similar to a feed-forward network. The neurons in the MLP are trained with the use of the back propagation learning technique. Since MLPs are designed to approximate any continuous function, they may be able to tackle problems that are not linearly separable. The primary application scenarios for machine learning and pattern recognition, classification, and approximation are presented.

1.3.6 Logistic Regression

Using prior observations from a data collection, a statistical analysis technique known as logistic regression forecasts a binary answer, such as yes or no [21]. A logistic regression model predicts a dependent data variable by looking at the association with one or more independent variables that are already available. For example, a logistic regression could be used to predict the outcome of an election for a public office or the acceptance rate of a high school student into a particular university. These straightforward decisions between two possibilities yield binary outcomes.A logistic regression model can be used to account for many input criteria.

1.3.7 K-Nearest Neighbor Classifier

K-Nearest Neighbor, sometimes known as KNN, is a supervised machine learning technique. It has already used in classification of brain stroke images [22]. The dataset is clustered in this algorithm. It uses the distance between data points to classify unclassified data. Euclidean, Manhattan, and Minkowski distances are possible distance measures. First, determine how many neighbors there are, or e.i. ”K”. Next, choose a non-classified data point. Next, determine the separation between the data point and its k nearest neighbors. Count the number of data points in the cluster among these k neighbors.At last assign the data point in that cluster which have maximum number of neighbor. If data point position (X1, Y1) and its neighbor position (X2, Y2) then Euclidean distance between them can be calculated by following equation:

$$d(x,y) = \sqrt{(X2 - X1)^2 + (Y2 - Y1)^2} \tag{5}$$

1.3.8 Convolutional Neural Network

Convolutional neural networks (CNNs) are deep learning network structures that learn directly from data. CNNs are very useful in recognizing patterns in images that depict objects [23]. Algorithms can be very helpful in classifying non-image data, such as audio, time series, and signal data. The traditional architecture of a CNN consists of convolutional, pooling, and fully connected layers.

The foundational component of the CNN is the convolution layer that performs the majority of the computation complexity on the network. This layer creates a dot product among two matrices, one being the kernel—a collection of hyper parameters the other of which is the constrained area ofthe input patch. Compared to a picture, the kernel is thinner in space but deeper which indicates that the kernel height and width will be relatively minimal if the picture comprises of 3 (RGB) channels, but the intensity will go upward to all three color channels. The size of the output volume may be determined using Equation 6 if such input is of size W x W x D, the numbers of kernels is Dout, and the dimension is F with stride S and amount of padding P.

$$W_{Out} = \frac{W - F + 2P}{S} + 1 \tag{6}$$

The pooling layer replaces the output of the network at certain points by computing an aggregate statistic from the surrounding results. This helps to reduce the dimension of the feature in the representation, which in turn reduces the amount of calculation time and weights required. The pooling process is applied independently to each section of the presentation. Let activation map of W x W x D, Stride S, Pooling Kernel Size F, output size can be calculated by the following equation:

$$W_{Out} = \frac{W - F}{S} + 1 \tag{7}$$

Every neuron in this layer is fully connected to every other neuron in the layer before and after, just like in a traditional FCNN. [24]

1.3.9 ***Proposed Ensemble Technique***

Figure 5denotes our proposed ensemble techniques. After training with seven different machine learning algorithms, we predict the validation set with each of the trained models weight and save those predictions into an array. For example, for a single image input from the validation set we get seven predictions from seven different algorithms. Let these predicted array are x1, x2, x3, x4, x5, x6 and x7. Now we concatenate these seven values into a 1D array so,

$$X = [X_1,\ X_2,\ X_3,\ X_4,\ X_4,\ X_5,\ X_6,\ X_7] \tag{8}$$

Than we concatenate this array with its true value which is present in the validation set. Let, true value is y1 and array is X2 Thus,

$$X2 = [X_1,\ X_2,\ X_3,\ X_4,\ X_4,\ X_5,\ X_6,\ X_7,\ Y1] \tag{9}$$

Likewise for each of the input image we get an 1D array and finally for the overall validation set we receive an 2D array that contains each of the images predictions for all 7 algorithms along with this true value. Thus, $X2$ = [[x_1, x_2, x_3, x_4, x_5, x_6, x_7, $Y1$], [x_1, x_2, x_3, x_4, x_5, x_6, x_7, $Y1$]......[x_1, x_2, x_3, x_4, x_5, x_6, x_7, $Y1$]].

Finally we transform this 2D array into a data frame. Table 1and Table 2demonstrates a sample of the MRI and CT scan dataset. We train our model using 7-bit binary system where,

$$2^7 or 128 binarydigit = \big[[0,0,0,0,0,0,0][0,0,0,0,0,0,1],[0,0,0,0,0,1,0],\ldots[1,1,1,1,1,1,1]\big] \tag{10}$$

Combination of 0s and 1s. We add another column using majority count of0S and 1S. If any row the number of 0S is greater than the number of 1S than the 8th column value of that row is 0 otherwise 1.Than we do raw shuffling of this binary values and convert it to the dataframe. And train any classifier by this data frame. Thus we create our ensemble model. We evaluate this model using our main data set that we calculated of equation (9).

Table 1:PROPOSED ENSEMBLE TECHNIQUE:NEWLY COMBAINED DATASET (CT SCAN IMAGES)

RF	DT	Naive Bayes	LR	MLP	XGBoost	CNN	Y1
0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0
0	0	1	1	1	1	1	1

RF	DT	Naive Bayes	LR	MLP	XGBoost	CNN	Y1
1	0	0	0	0	0	0	0
0	0	1	1	1	1	1	1
0	1	0	1	1	1	1	1

Table 2: PROPOSED ENSEMBLE TECHNIQUE: NEWLY COMBAINED DATASET (MRI IMAGES).

CNN	KNN	LR	XGBoost	MLP	Naive Bayes	DT	Y1
1	0	1	1	1	1	1	1
1	0	0	0	1	1	0	0
0	0	0	1	1	1	1	1
0	0	0	0	0	1	1	0
1	1	0	1	0	0	0	0
1	1	1	1	1	0	0	1

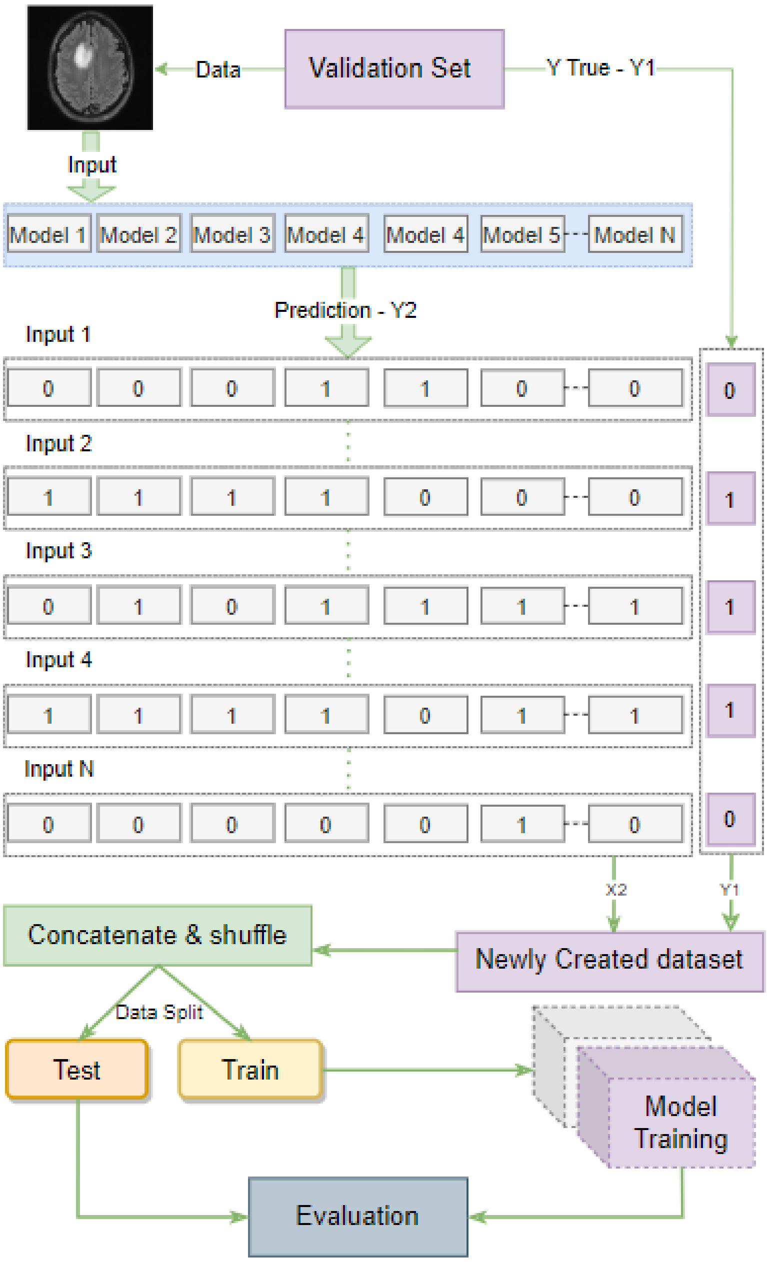

Figure 5: Proposed Ensemble Model.

IV. Performance Evaluation Metrics and Results

1.4 Evaluation Metrics

The degree to which a system's predictions agree with the actual data is a measure of its accuracy. It is expressed as a percentage of all forecasts that the framework accurately predicts. Accuracy becomes more important when true positives and true negatives are more important than false positives and false negatives. In this study, an instance of a true positive occurs when our system correctly identifies anomalies, whereas an instance of a false positive occurs when our system is unable to detect irregularities.

$$Accuracy = \frac{Correct\ \ Predictions}{All\ Prediction\ \ s} \tag{11}$$

The equation 11 provides the procedure to determine a system's performance. The number of correct forecasts is divided by the total number of predictions in this case. The number of times a model achieved the correct forecast out of all the predictions it made is represented by its accuracy. When the prediction accuracy is 88%, for instance, it indicates that out of 100 forecasts, 88 came true for the model.

Real positives and false positives can be distinguished using Precision. A false positive in our system happens when it makes an incorrect prediction about what anomaly it has detected.

		Predicted	
		Negative	Positive
Actual	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

A model's precision can be measured by how accurate or precise each and every one of its positive predictions was. It might also be used to determine the accuracy with which the algorithm predicts negative values when the equation is utilized 12.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{12}$$

Recall, as precision, is based on relevancy. The proportion of right predicted out of all the accurate predictions which should have been made is known as recall.

$$Recall = \frac{True\ Positive}{Total\ Actual\ Positive} \tag{13}$$

As par equation 13, the percentage of accurate predictions divided by the total number of correct predictions it should have made is known as recall. When a model has a recall value of 70%, for instance, it means that 70 of the 100 correct predictions it should have produced were made by the model.

The F1 score is not commonly utilized as an indicator of accuracy, precision, or recall, but it does offer a decent balance of precision and recall. The accuracy and recall of a test are used to compute the F1 score. According to the equation 14, the F1 score is more relevant than accuracy if false negatives and false positives play a substantial role.

$$F1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{14}$$

The value of the F1 score becomes 1 when precision and recall are perfect. This is also the maximum F1 score that may be achieved. An F1 score of 0 is the worst-case situation, which happens when either the accuracy or recall is zero.

1.5 Performance Evaluation

As previously stated, we have employed seven different machine learning model in the initial stage that are Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, Logistic Regression, Multilayer Perceptron, Extreme Gradient Boosting and Convolutional Neural Network. In order to evaluate our models, we have compared between a series of performance metrics.

Figure 6depicts the comparison of these employed performance metrics that are AUC ROC, precision, recall, F-1 score. Here, we can visualize that most of the model’s metric values are stable with close precision, recall, F-1 score.On the other hand, a lower recall and f-1 score are associated with agreater precision value in Gaussian Naive Bayes, Decision Tree, and Random Forest classifiers.While predicting critical decisions such as brain stroke true positive and false negative values play a very crucial role. Thus if implemented directly, these three discussed model’s prediction can have very severe consequences.

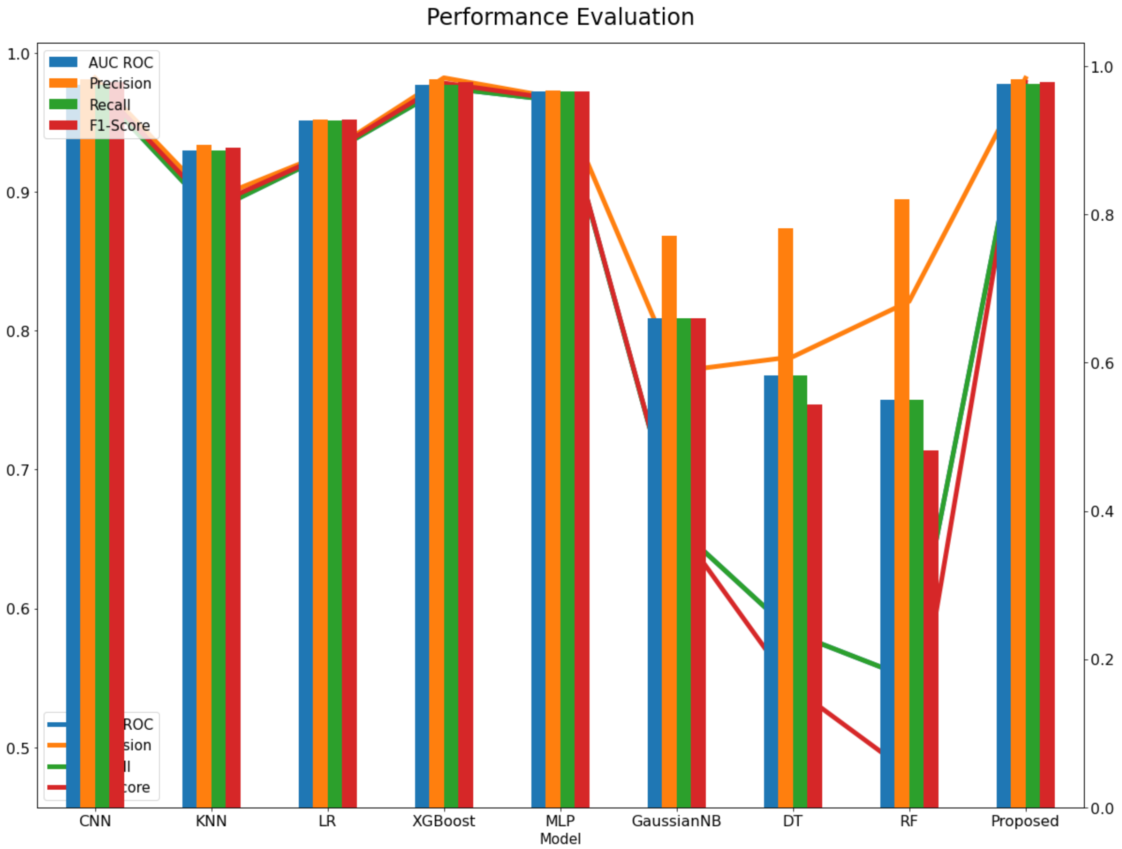

Figure 6: System Performance: Performance Metrics Comparison.

Figure 7 reflects the accuracy score gained by each of our trained algorithm. Here, we can see except Gaussian Naive Bayes, Decision Tree and Random Forest almost all of the algorithms managed to attain a satisfied score. Moreover, our proposed model managed to suppress all of the utilized machine learning algorithms.

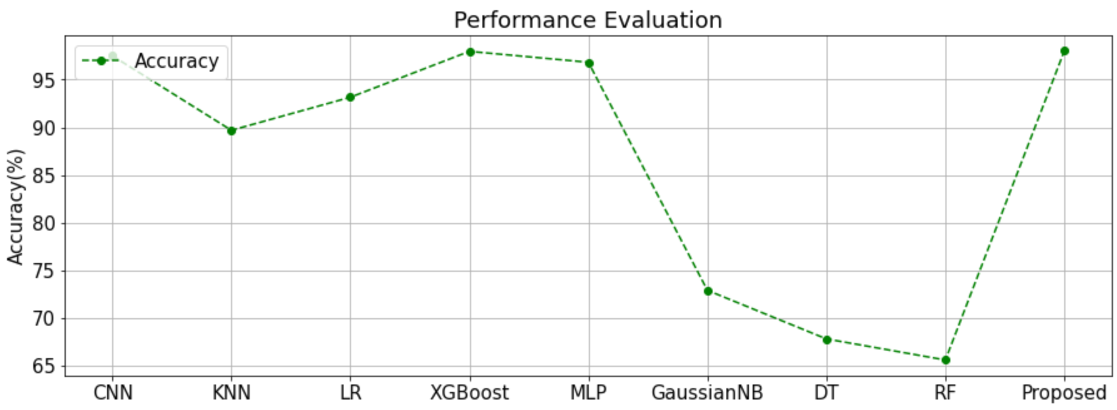

Figure 7: System Performance: Accuracy Comparison

Table 3denotes the classification report of both classes for each of the trained models. Here, it is indeed clear that random forest's precision, recall and f-1 score is satisfactory in detecting normal images. Yet, on stroke data, its recall and f-1 score almost vanishes meaning our trained model is biased towards normal cases and most likely to predict Normal. Thus, our model will have higher False Negative values. Decision Tree also acts in a similar way. Although the situation improves in (d)Gaussian Naive Bayes, the result is still not satisfactory. However, the rest of the algorithm's precision, recall and f-1 score in both classes are satisfactory. In terms of both normal and stroke cases our proposed system managed to reach almost 98%-99% in precision,recall and f-1 score. This values states that our model is not biased towards any particular class and can identify between stroke and normal cases almost flawlessly.

Table 3: BRAIN STROKE CLASSIFIER: CLASSIFICATION REPORT COMPARISON

Model	Class	Precision	Recall	F-1 Score
Random Forest	Normal	0.64	1.00	0.78
	Stroke	1.00	0.10	0.18
Decision Tree	Normal	0.66	0.99	0.79
	Stroke	0.90	0.18	0.30
Gaussian Naive Bayes	Normal	0.71	0.95	0.81
	Stroke	0.83	0.36	0.51
KNN	Normal	0.90	0.93	0.92
	Stroke	0.89	0.84	0.86
Logistic Regression	Normal	0.94	0.95	0.94
	Stroke	0.91	0.91	0.91
MLP	Normal	0.97	0.98	0.97
	Stroke	0.96	0.95	0.96
XGBoost	Normal	0.97	0.99	0.98
	Stroke	0.99	0.96	0.97
CNN	Normal	0.99	0.99	0.99
	Stroke	0.99	0.98	0.98
Proposed System	Normal	0.98	0.99	0.98
	Stroke	0.99	0.96	0.97

Moreover, Figure 8signifies the confusion matrix for all eight of the trained models. Here, we can visualize that although (e)Random Forest and (f)Decision Tree has a good number of true positive and true negative values, these model fails to detect stroke cases and has a huge number of False Negatives. Naive Bayes also has the same characteristics. In (a)logistic regression the situation improved drastically. Also, (b)Extreme Gradient Boosting, (c)MLP and (g)CNN manages to score more true positive and true negative values. Finally, in our (h)proposed model, we only have 42 false negative and 13 false positive values that proves that our model can almost seamlessly classify between stroke and normal images.

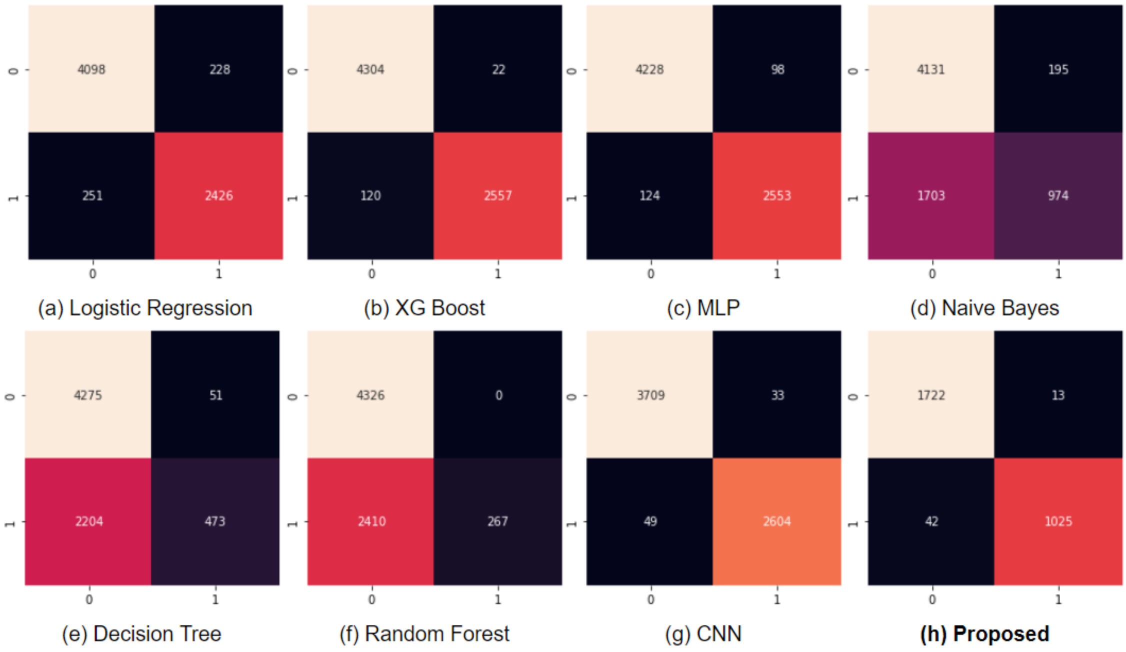

Figure 8: System Performance: Confusion Matrix Comparison.

Finally, Table 4 and Table 5reflects our overall findings. As stated previously, although Random Forest, Decision Tree and Gaussian Naive Bayes managed to score good accuracy these algorithms fails to identify stroke classes and is biases toward normal cases. Thus, these model has a very low AUC, recall and f-1 score. Suppressing these three mentioned algorithm Logistic Regression reach an accuracy of 93.16%. Moreover, it has a AUC score of 92.676% and very similar average recall, precision and f-1 score. The remaining algorithms Multilayer Perceptron, Extreme Gradient Boosting, Convolutional Neural Network managed to gain accuracy around 96-97%with satisfactory AUC, average precision, recall, f-1 scores. Our proposed ensemble method however, reached an accuracy of 98.37% along with 97.66% AUC, 98.18% precision, 97.66% recall and 97.91% f-1score which proves our system to be more efficient than the utilized machine learning models. It also be noted that our model give 98% accuracy in CT scan dataset but it gives 94% accuracy using MRI images dataset. So this model gives better accuracy by using CT images comparison to MRI images.

Table 4: BRAIN STROKE CLASSIFIER: PERFORMANCE METRICS COMPARISON (FOR MRI IMAGES DATASET)

Model	**Accuracy**	**Loss**	**Recall**	**Specificity**	**Precision**	**F-1 Score**
Convolutional Neural Network	91.76	0.0478	0.9033	0.9406	0.9444	0.9201
K-Nearest Neighbor Classifier	77.52	0.2248	0.5851	0.9482	0.9112	0.7126
Logistic Regression	72.48	0.2752	0.7191	0.7299	0.7077	0.7134
XGBoost	90.29	0.0971	0.8732	0.9299	0.9189	0.8955
MLP	80.28	0.1972	0.744	0.8562	0.8247	0.7823
Gaussian Naive Bayes	66.13	0.3387	0.5778	0.7372	0.6667	0.6191
Decision Tree Classifier	66.06	0.3394	0.8034	0.5307	0.6089	0.6927
Proposed Ensemble Method (Using Support Vector Classifier)	94.06	0.0594	0.9121	0.967	0.9625	0.9366
Proposed Ensemble Method (Using K-Nearest Neighbor Classifier)	94.36	0.0564	0.9376	0.9491	0.9448	0.9412

Table 5: BRAIN STROKE CLASSIFIER: PERFORMANCE METRICS COMPARISON (FOR CT SCAN IMAGES DATASET)

Model	Accuracy	Loss	AUC	Precision	Recall	F-1 Score
Random Forest Classifier	65.586	0.1294	0.54986	0.82111	0.5498	0.481761
Decision Tree Classifier	67.7995	0.1261	0.58245	0.78124	0.5824	0.5434
Gaussian Naive Bayes	72.8973	0.1035	0.65938	0.77064	0.65938	0.6598
Logistic Regression	93.160	0.0832	0.92676	0.92818	0.9267	0.92746
MLP	96.8299	0.0678	0.96551	0.96727	0.96551	0.96637
XGBoost	97.9722	0.0502	0.975044	0.98217	0.97504	0.97837
Convolutional Neural Network	98.72	0.0574	0.98635	0.98722	0.98635	0.98678
Proposed Ensemble Method	98.0371	0.0402	0.97657	0.98183	0.97657	0.97907

V. CONCLUSION

Brain stroke is a fatal medical condition that requires precise diagnosis. Any misclassification of this condition might have potentially horrific consequences. Few research have been done in the past to identify brain strokes [25]; the majority of these investigations concentrate on patient health characteristics. Few research have focused on accurately diagnosing brain strokes using CT or MRI images, and none of the investigations have been successful. Additionally, the author only performs a single round of training in the existing studies, and in the event that the model fails to accurately classify, there are no alternative fallbacks. Even in event of misclassification it is crucial to identify if a prediction is a false positive or false negative. In this paper, we provide a cutting-edge ensemble method that makes utilization two stages of training. Our model initially derives from seven distinct machine learning methods. The output of those seven models is being used to produce a fresh dataset in the concluding phase. Finally, this new dataset is combined with true values and segmented once more for additional training. In this stage we further employ Extreme Gradient Boosting. Our suggested system demonstrated effectiveness and was able to overcome any machine learning algorithm's performance that we have employed in the earlier phase. With our proposed system we have managed to reach 98.0731% with satisfied AUC, precision, recall, f-1 values thus proving our model can identify both of the classes almost precisely and not biased towards any group. In addition, the confusion matrix reveals that among 2852 cases our system can identify 2747 events correctly.

ACKNOWLEDGMENT

First author, Alhina Afrin Dolon, is a research fellow of the Ministry of Post, Telecommunication and Information Technology of the Government of Bangladesh (GoB). She obtained that fellowship as a master student of Computer Science and Engineering Department of JatiyaKabiKaziNazrul Islam University, Bangladesh. Therefore, we happily acknowledge that support and we would like to devote a thank to the stated ministry of GoB.

REFERENCES

[1] *Tudor G Lu HT Kadirvel R Zhou G, Li MH and Kallmes D. Remote ischemic conditioning in cerebral diseases and neurointerventional procedures: Recent research progress. Frontiers in Neurology., 9:339,*

2018..

[2] *World Health Organization. Who reveals leading causes of death and disability worldwide: 2000-2019, 12 2020..*

[3] *Norrving B et al. Feigin VL, Brainin M. World stroke organization (wso): Global stroke fact sheet 2022. International Journal of Stroke., 17(1):18–29, 2022.*

[4] "Rita V Krishnamurthi, Takayoshi Ikeda, and Valery L Feigin. Global, regional and country-specific burden of ischaemic stroke, intracerebral haemorrhage and subarachnoid haemorrhage: a systematic analysis of the global burden of disease study 2017.," *Neuroepidemiology*, vol. 54(2), pp. 171-179, 2020.

[5] *Emelia J Benjamin, Michael J Blaha, Stephanie E Chiuve, Mary Cushman, Sandeep R Das, Rajat Deo, Sarah D De Ferranti, James Floyd, Myriam Fornage, Cathleen Gillespie, et al. Heart disease and stroke statistics—2017 update: a report from the american heart,* association circulation, 135(10):e146–e603, 2017.

[6] *Antonio Di Carlo. Human and economic burden of stroke, 2009.*

[7] *Ailton Andrade de Oliveira, Maria Teresa Carthery-Goulart, Pedro Paulo de Magalh˜aes Oliveira J´unior, Daniel Carneiro Carrettiero, and Joao Ricardo Sato. Defining multivariate normative rules for healthy aging using neuroimaging and machine learning:,* an application to alzheimer's disease. Journal of Alzheimer's disease: JAD, 43(1):201– 212, 2015.

[8] *Deanna Greenstein, James D Malley, Brian Weisinger, Liv Clasen, and Nitin Gogtay. Using multivariate machine learning methods and structural mri to classify childhood onset schizophrenia and healthy controls. Frontiers in psychiatry, 3:53, 2012..*

[9] *Maya Bleich-Cohen, Shahar Jamshy, Haggai Sharon, Ronit Weizman, Nathan Intrator, Michael Poyurovsky, and Talma Hendler. Machine learning fmri classifier delineates subgroups of schizophrenia patients. Schizophrenia research, 160(1-3):196–200, 2014.*

[10] *Jie An, Peng Fang, Wensheng Wang, Zhenyin Liu, Dewen Hu, and Shijun Qiu. Decreased white matter integrity in mesial temporal lobe epilepsy: a machine learning approach. Neuroreport, 25(10):788–794, 2014.*

[11] *Negin Moghim and David W Corne. Predicting epileptic seizures in advance. PloS one, 9(6):e99334, 2014.*

[12] *Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation.,* Medical image analysis, 36:61–78, 2017.

[13] *Ching-Heng Lin, Kai-Cheng Hsu, Kory R Johnson, Yang C Fann, ChonHaw Tsai, YU Sun, Li-Ming Lien, Wei-Lun Chang, Po-Lin Chen, ChengLi Lin, et al. Evaluation of machine learning methods to stroke outcome prediction using a nationwide disease registry.,* Computer methods and programs in biomedicine, 190:105381, 2020.

[14] *JoonNyung Heo, Jihoon G Yoon, Hyungjong Park, Young Dae Kim, Hyo Suk Nam, and Ji Hoe Heo. Machine learning–based model for prediction of outcomes in acute stroke. Stroke, 50(5):1263–1265, 2019.*

[15] *Adam Hilbert, Lucas Alexandre Ramos, Hendrikus JA van Os, S´ılvia D Olabarriaga, Manon L Tolhuisen, Marieke JH Wermer, Renan Sales Barros, Irene van der Schaaf, Diederik Dippel, YBWEM Roos, et al.,* Dataefficient deep learning of radiological image data for outcome prediction after endovascular treatment of patients with acute ischemic stroke. Computers in biology and medicine, 115:103516, 2019.

[16] *Burak Tasci and Irem Tasci. Deep feature extraction based brain image classification model using preprocessed images: Pdrnet. Biomedical Signal Processing and Control, 78:103948, 2022.*

[17] *Brain stroke prediction ct scan image dataset.*

[18] *Tin Kam Ho. Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition, volume 1, pages 278–282. IEEE, 1995.*

[19] *Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, et al. Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4):1–4, 2015.*

[20] *Daniel Svozil, Vladim´ır Kvasnicka, and Jir´ı Pospichal. Introduction to multi-layer feed-forward neural networks. Chemometrics and Intelligent Laboratory Systems, 39(1):43–62, 1997.*

[21] *Raymond E Wright. Logistic regression. 1995.*

[22] *Soundarapandian R.K. Gandomi A.H. et al. Govindarajan, P. Classification of stroke disease using machine learning algorithms. Neural Comput Applic, 32:817–828, 2020.*

[23] *Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. Recent advances in convolutional neural networks. Pattern recognition, 77:354–377, 2018.*

[24] *Ilkay Oksuz. Brain mri artefact detection and correction using convolutional neural networks. Computer Methods and Programs in Biomedicine, 199:105909, 2021.*

[25] *Manisha Sanjay Sirsat, Eduardo Ferm´e, and Joana Cˆamara. Machine learning for brain stroke: A review. Journal of Stroke and Cerebrovascular Diseases, 29(10):105162, 2020.*